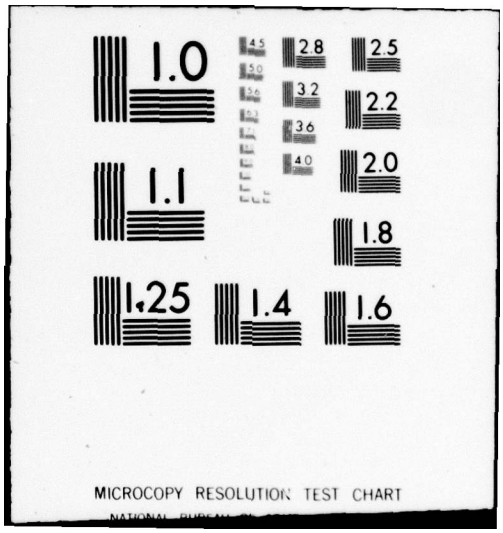
AD-A074 102 WALTER REED ARMY INST OF RESEARCH WASHINGTON DC F/G 12/1
SOME RANK-ORDER TESTS FOR TREND IN A SET OF CORRELATED MEANS,(U)
JUN 57 A LUBIN

UNCLASSIFIED NL

| OF |
ADA
074102

END
DATE
FILMED
10 —79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART

Some Rank-Order Tests for Trend in a Set of Correlated Means,

(10) Ardie Lubin

Walter Reed Army Institute of Research

Washington 12, D. C.

In many experiments the major interest is not in the amount of difference caused by the experiment, but the rank-order which results. This is especially true when n successive measurements are made on the same m subjects, where the subjects are being exposed to increasing amounts of work, sleep loss, etc. For such studies, the null hypothesis is that there is no trend; the usual alternative hypothesis is that the means are a monotonic function of time.

The motivation for this paper comes mainly from A. R. Jonckheere's recent development of a general non-parametric test against ordered alternatives which can be used to test the hypothesis that a set of correlated means has a predicted rank-order (1954a, 1954b). Jonckheere uses a statistic P (based on the Kendall rank-order correlation, tau) which is the sum of Kendall's P_i values, computed between the predicted rank-order and the observed rank-order for each of the m subjects.

This paper will propose an alternative statistic J (based on Spearman's rank-order correlation, rho) which can be shown to be more powerful than P in some special cases. J is the sum of the $S_i(d^2)$ values computed between the observed ranking of the n scores for each subject and the hypothetical ranking of the n scores; i.e.,

DEPARTMENT OF THE ARMY
ARI FIELD UNIT, BENNING
U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
P.O. BOX 2086, FORT BENNING, GEORGIA 31905

PERI-IJ

8 August 1979

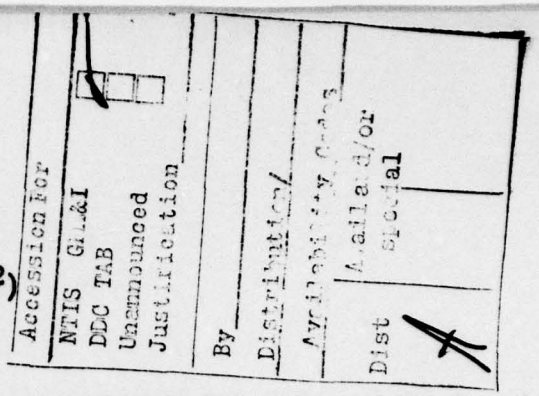
SUBJECT: Shipment of Documents

Defense Documentation Center
Cameron Station
Alexandria, VA 22314
ATTN: Selection & Cataloging

The Documents in these shipments are approved for public release. The distribution is unlimited.

FOR THE CHIEF:

ALEXANDER NICOLINI
Major, Infantry
R&D Coordinator



$$(1) \qquad J = \sum_{i=1}^{m} S_i(d^2)$$

The assumptions underlying such rank-order tests as the P and J procedures will be discussed in detail and a robust analog of J will be proposed. Some experimental designs will be suggested for which these trend tests are appropriate.

J is related in a rather simple way to other non-parametric tests. It has already been noted that J is simply the Spearman rank-order analog of P. Since Spearman's rho is more sensitive than Kendall's tau, J is more sensitive, and sometimes more powerful, than P. (The relative power of the two tests will be discussed in detail later.) On the other hand, for small values of n, P can be approximated by the normal distribution much better than J.

J is exactly equivalent to the one-tail binomial sign test, when $n = 2$. So the J test can be regarded as an extension of the sign test to the case where $n > 2$.

The average rank-order correlation between the m rankings and the predicted rank-order is given by

$$(2) \qquad K = 1 - \frac{6J}{m(n^3-n)} .$$

K can be regarded as a special case of Kendall's W coefficients: W is a linear function of the average Spearman rank-order intercorrelation of m rankings with one another, whereas K is the average correlation of m rankings with a hypothetical rank-order.

If two assumptions are made, it can be shown that J is normally

distributed for large values of m or n.

Assumption 1: For any subject, all permutations of the n socres are equally likely.

Assumption 2: The rank-order for any subject is statistically independent of the rank-order for any other subject.

Kendall (1948) has shown that Assumption 1 leads to a normal distribution of $S_i(d^2)$ when n is large. Therefore when n is large, J is normally distributed, since the sum of normally distributed variables is itself normally distributed.

Assumption 2 leads to a normal distribution for J when m is large, since, by the well-known Central Limit Theorem, a sum of m independent random variables will tend to normality as m increases.

M. G. Kendall (1948) gives the mean of $S_i(d^2)$ as

$$\frac{1}{6}(n^3-n), \tag{3}$$

and the variance as

$$\left(\frac{n^3-n}{6}\right)^2\left(\frac{1}{n-1}\right). \tag{4}$$

Since the mean of a sum is equal to the sum of the means, the mean of J is

$$\mu = \frac{m}{6}(n^3-n). \tag{5}$$

Since the variance of a sum of uncorrelated variables is equal to the sum of the variances it follows that the variance of J is

$$(6) \qquad \sigma^2 = \frac{mn^2(n+1)^2(n-1)}{36}$$

The value of $S_i(d^2)$ is always even, so the interval between adjacent values of J is always two. Therefore the correction for continuity is unity.

From the above, it follows that for large values of m and n,

$$(7) \qquad z = \frac{J-\mu+1}{\sigma}$$

is normally distributed, with a mean of zero and a standard deviation of unity, when the null hypothesis is true. Since the alternative hypothesis is directional, only negative values of z need be tested for significance; the null hypothesis is accepted automatically if z is positive or zero. The large sample test in terms of K is

$$(8) \qquad z = \left[K - \frac{6}{m(n^3-n)}\right]\sqrt{m(n-1)}$$

Since it is a one-tail test, only positive values of z need be tested for significance; the null hypothesis is accepted automatically if z is negative or zero.

The distribution of J for small values of m and n tends to normality much more slowly than does that of P. The distribution of $S_i(d^2)$ is symmetric but, as Kendall (1948) has pointed out, has the unusual property that values close to the mean are not necessarily more frequent than values further from the mean. This gives the distribution curve a peculiar sawlike profile. Since these reversals occur mostly near the mean, the normal curve gives a better fit to the tails of the cumulative

distribution than it does near the mean. Fortunately, in tests of significance, we are mainly interested in the fit near the tails.

In general, J reflects these properties of $S_i(d^2)$ although the reversals disappear rather quickly as m increases. Table 7 gives the probability distribution for J where the normal curve gives a poor fit to the tail of the distribution. Let us arbitrarily define the tail of a distribution as that portion where the cumulative probability is less than .100. Then, as a rough guide, we can say that when mn is greater than 12, the maximum error in the distribution tail will be .004 or less. The maximum error near the mean, when $mn > 12$, is .009 or less. Distribution tables are given in this article for all cases where the maximum tail error is .004 or greater, except where $m = 1$ or $n = 2$. When $n = 2$, J is, of course, the binomial sign test; when $m = 1$, J becomes equal to $S_i(d^2)$. Adequate tables for these cases can be found elsewhere (e.g., E. S. Pearson and H. O. Hartley, 1954, pp. 186, 211 and 238).

To fix our ideas let us apply the J test to a learning example used by Jonckheere.

Insert Table 1 about here

In Table 2, the rank-order scores for the values in Table 1 are given. The null hypothesis is that there is no trend. The alternative

---

Insert Table 2 about here

---

hypothesis is that the number of errors decrease from the first learning trial to the seventh.

Here there are five subjects and seven scores, so m = 5 and n = 7. Since mn > 12, the large sample test will be used. The expected mean and standard deviation are

$$\mu = \frac{m(n^3-n)}{6} = 280,$$

$$\sigma = \frac{n(n+1)}{6}\sqrt{m(n+1)} = 51.121.$$

The normal deviate is therefore

$$z = \frac{J-\mu+1}{\sigma} = -2.84 .$$

Thus the result is significant beyond the .01 confidence level. If z had been positive, the null hypothesis would have been accepted.

We can also calculate the normal deviate for Jonckheere's P test. The mean of P is mn(n-1)/4; the variance is mn(n-1)(2n+5)/72; and the correction for continuity is 1/2. (The sign of P is always opposite to that of J.)

In this case J is more powerful than P. As will be shown later, this is because of its greater sensitivity.

Now let us apply the J test to a case where m and n are small. In Table 3 a fictitious example is given where the number of seconds, of alpha EEG wave frequency during a 60-second interval, has been recorded for each subject during four successive days of sleep loss. In Table 4, the rank-order values are given. The null hypothesis is that there is

---

Insert Tables 3 and 4 about here

---

no trend. The alternative hypothesis is that EEG waves of the alpha type appear less frequently as sleep loss increases.

Since there are three subjects and four scores, mn = 12. From Table 7 we find that an observed value of $J \leqq 12$ (for m = 4 and m = 3) will occur with a probability of .045, so the average Spearman rank-order coefficient of .60 is significant at the .05 level.

An mn value of 12 is just within the arbitrary borderline we have drawn for small samples. How much error would there be if the normal approximation were used?

$$z = \frac{J-\mu+1}{\sigma} = -1.70$$

This z corresponds to a one-tail probability of .0446 as compared to the exact probability of .045. Clearly, the normal approximation is nearly perfect here.

Using Jonckheere's trend test,

$$z = \frac{P-\frac{1}{2}-\frac{1}{4}mn(n-1)}{\sqrt{\frac{mn}{72}(n-1)(2n+5)}} = \frac{12-9.5}{2.550} = .980$$

Thus, using the P test in this case would have led to the acceptance of the null hypothesis. This example was deliberately constructed to show that a large difference can occur between the J test and the P test. (Later we will construct an example where P is more powerful than J.)

Alternatives to the J Test

The rank-order tests for trend seem to have no parametric analogs. If we are willing to specify the exact differences that theoretically should exist between the n correlated means, Hotelling's T test (Hotelling, 1931) enables us to compute the likelihood that the observed set of n means could be a sample from the theoretical universe. But this in general places too great a burden upon the experimenter. Ordinarily, he can only specify the direction in which the subject's scores should move, not the amounts. So there is no multivariate normal statistic which is exactly analogous to the J test. A series of t-tests could be used, but the significance level would be hard to determine for any given sequence of results.

What might be called a quasi-parametric test of trend has been proposed (personal communication) by N. Mantel of the National Institutes of Health.

Let Y_{ij} be the score of the i^{th} subject on the j^{th} treatment, and

X_j be the theoretical rank of the j^{th} treatment.

Then; b_i, the slope of Y on X, can be calculated in the usual least-squares manner for the i^{th} subject. If the null hypothesis is correct, the average value of b is zero; if the alternative hypothesis is correct, the average is positive.

When the distribution of observed b's is normal, a one-tail t test with (m-1)d.f is most powerful, otherwise a permutation or sign test can

be used. The fundamental assumptions here are: (1) X is a positive linear function of X, (2) the n scores for a subject are statistically independent of the scores for any other subject.

Clearly, the first assumption is not absolutely necessary. If the theoretical trend is any monotonic function, Mantel's test will be useful because it allows us to compare the strength of trends. The non-parametric tests, J and P, are tests of consistency rather than amount of change. Suppose we are trying to compare the effect of sleep loss on simple and disjunctive reaction time. Both measures may give us a K value of unity, and yet the increase in reaction time could be considerably greater for one task.

Another alternative which preserves the metric is R. A. Fisher's permutation test. Here it is assumed that for any subject, his set of n observed scores are purely a chance arrangement and any other permutation of the n observed scores is equally likely. This is clearly Assumption 1 with the addition that the set of observed scores is an adequate representation of the universe. In general Fisher's permutation procedure requires the computation of $(n!)^m$ sets of n correlated means. The J and P tests may be viewed as permutation tests on rank-order transformations of the raw scores.

Presumably there are other non-parametric trend tests that could be used in place of J or P. For example, the set of n correlated means could be rank-ordered and a Kendall tau or Spearman rho computed. Essentially, this is a rank-order trend test where m is always unity. As Jonckheere has pointed out, such a test would always be less powerful than the J and

P tests which take account of the size of m.

We have several times previously noted that the J test seems in some cases to be more powerful than the P test. This results from the fact that Spearman's $S_i(d^2)$ is usually a more sensitive measure of rank-order correlation than Kendall's P. For example, take the rank-orders 1342, 1423, 2314, 3124, and 2143. All of these have the same Kendall P value, 4. If we compute Spearman's $S_i(d^2)$, then four of these rank-orders end up with the same value, 6; but the rank-order 2143 has an $S_i(d^2)$ of 4 (indicating a Spearman rho of .60 compared with .40 for the other rank-orders). It is these discriminations which cause the sawtooth profile for the distribution of rho.

The relation between Spearman's rho and Kendall's tau is rather complex. When n = 2 or 3, they are identical (except for a linear transformation). When n = 4, tau is a single-valued function of rho, but rho is not a single-valued function of tau. Therefore, when n = 4, rho is more powerful than tau. When n is greater than 4, it is still true that there are many more possible values of rho than tau, but tau is no longer a single-valued function of rho. There are some relatively infrequent occasions when there are several values of tau for each value of rho. So, although rho will be more powerful than tau for most non-null universes than can be constructed, it is not uniformly more powerful. (I am grateful to M. P. Schutzenberger of the Massachusetts Institute of Technology for his help in working out this relation.)

This means that it is possible to construct a non-null universe where P will be more powerful than J. For example, when n = 5, take a

population where the rank-order is always 23451. The Spearman rho equals zero, whereas the same rank-order gives a value of .20 for Kendall's tau. Clearly, J will never be significant no matter how large m becomes whereas P will be significant at the 5 percent level for $m \geqq 13$.

[It is of some interest to ascertain whether a statistic such as $S(d^q)$, where q is even, might prove more powerful than either $S(d^2)$ or P. It turns out that $S(d^q)$ gives results equivalent to $S(d^2)$ for $n = 2, 3, 4$, and 5, but does introduce additional discrimination for $n > 5$. The additional discrimination seems to be slight and this approach has not been pursued further.]

In general then, there is no exact parametric analog for the J test, and J will usually (but not always) be more powerful than P.

Appropriateness of the J Test

Under what circumstances should the J test be applied? To answer this question, let us re-examine the two basic assumptions and consider what alternative hypotheses we want to test.

(1) For any subject, all permutations of the n scores are equally likely.

(2) The rank-order for any subject is statistically independent of the rank-order for any other subject.

Assumption 2 need not trouble us. With the exception of such obvious cases as siblings, or matched groups, one subject's rank-order can always be expected to be independent of any other subject's rank-order if the null hypothesis is true and the n treatments have no effect.

But Assumption 1, though simply stated, raises a considerable number

of difficulties, the same ones that arise in using analysis-of-variance for a successive measurement design. These difficulties are of two sorts, experimental and statistical.

Experimental difficulties occur if there is any systematic result of the measurement itself (such as a practice effect). (E.g., if there is learning or fatigue, then even when the n treatments do not differ, there will be a trend in the data.) From the experimental design point of view, it is clear that somehow the "carry-over" effect of learning, fatigue, etc., must be dealt with before the effect of the treatments can be assessed. Two common designs for accomplishing this are the "plateau" procedure and the use of matched controls.

The first procedure involves testing the subject until his scores level off and a steady state is reached. There are several objections to this: (1) considerable time is required, (2) many plateaus can be encountered in the learning curve of one subject and it is difficult if not impossible to state when the ultimate level has been reached, (3) sometimes it is exactly the effect of the treatments on the speed of learning that we wish to ascertain.

The "matched control" design uses control subjects, paired with each experimental subject, who experience all of the n successive measurements without the treatments. A difference score, $D_i = E_i - C_i$, can then be computed for each pair on the i^{th} measurement. The expected value of this difference score will be constant if the treatments have no effect on the scores of the experimental subject.

A major disadvantage of this design is that it assumes the control

will have the same shape of learning (or fatigue) curve that the experimental subject would have had. If the matching is poor, considerable error variance can be introduced via this assumption. However, such errors will not bias the trend test in the population.

It is customary to use what are called "balanced" designs. Thus, if there are two treatments A and B, Group 1 would receive the treatments in the order AB; Group 2 would have the order BA. In general, the experimental design and assumptions are those of the Latin square. In particular, it is assumed that there is no interaction between treatments and order. Suppose, however, that under Treatment A, scores are depressed initially, but learning proceeds much more swiftly and a higher ultimate level is reached than under Treatment B. Then if a great many trials are given for each treatment, Treatment A will be judged to be superior. By shortening the number of trials, the advantage for A can be wiped out, and even reversed. Clearly, any interaction between treatment and order can lead to bias. And the very existence of effects such as learning, boredom, fatigue, etc., makes it unsafe to assume that there will be no interaction between such effects and the treatment.

Still another design difficulty arises if any of the treatments has residual effect; e.g., some drugs have a physiological effect that lasts for days and sometimes weeks. One usual method is to allow sufficient time to elapse between treatments so that the residual effects have been eliminated. Again, if the residual effects cannot be eliminated, balanced designs are of doubtful value. In general, psychological

experiments cannot meet the requirements that the treatment effects be constant regardless of order and that the residual effects be zero.

Unless the treatments are the practice effects of the successive measurements themselves (as in Jonckheere's example), "matched control" design is strongly recommended for use whenever successive measurements are involved.

Unfortunately, this still does not dispose of Assumption 2. Clearly, Assumption 2 involves statistical restrictions on the raw score distributions. For example, if one distribution is rectangular and the other is skewed, Assumption 2 does not hold. What conditions must be placed on the n-variate distribution of raw scores such that, when the n treatments have no effect, the assumption that "all permutations are equally likely" holds true?

Lehmann and Stein (1949) mention a particular kind of n-variate symmetry in which the univariate distributions of the n measurements are identical, the bivariate distributions of any two treatment variables are identical, the trivariate distributions of any three treatment variables are identical, etc. For an n-variate normal distribution this implies equal means, equal variances, and a constant intercorrelation. If any multivariate distribution of raw scores exhibits this symmetry, then of necessity the ELP (equally-likely-permutations-assumption) will hold. (I am indebted to Professor W. A. Wallis of the University of Chicago and Doctor J. R. Rosenblatt of the National Bureau of Standards for calling my attention to this particular kind of "multivariate" symmetry and for aid in interpreting the results.)

E. L. Lehmann (personal communication) has pointed out that although this particular multivariate symmetry is sufficient for the ELP, it is not necessary. For example, if we take any two non-identical independent distributions that are symmetric about the same mean, the ELP will hold, even though the univariate distributions are different. (Unfortunately, this doesn't extend to the case of three or more symmetric distributions.) So far as I know, a set of conditions which would be both necessary and sufficient for the ELP has not been devised.

This requirement of the ELP is a very severe restriction upon the use of the J test. Essentially, J is a test of the homogeneity of the treatment variables. If J is significant, it is assumed that this is due to differences between the n means, but the significance of J may be due to differences in the variance, third moment, fourth moment, etc. Any deviation from "multivariate" symmetry could conceivably cause a significant J. This difficulty is not confined to the J test but seems to apply to all rank-order tests for differences between means; e.g., Kendall's W, Kruskal-Wallis H, Festinger's d, etc. The demand that all treatment variables have a constant dependency upon one another is strikingly similar to the demand in two-way analysis of variance, that the treatment variables have a constant intercorrelation.

What can be done if one suspects that the ELP does not hold? At least two different paths are possible. One solution is a procedure to test whether a significant J is due to the differences between the n means or to the deviations from ELP. The other way is to construct a robust test which does not demand ELP.

A Preliminary Test for J

Let us take the raw scores for each treatment and convert them into deviations from the treatment mean. Then the mean of the deviate scores for each treatment is zero. Now rank-order the deviate scores for each subject in the usual way and perform the J test. If J is significant, it must be due to deviations from the ELP since the treatment means are equal. For convenience I shall call this preliminary procedure the A test.

Let us apply the A test to Jonckheere's example. This has been done in Tables 5 and 6. Since the z from Table 6 is positive and less than

---

Insert Tables 5 and 6 about here

---

unity, we conclude that the significant z found from the data in Table 1 is not due to deviations from ELP.

A Robust Analog of J

An alternative to a preliminary test would be to alter the J test so that it would be robust to deviations from the ELP. (I am indebted to Doctor S. Greenhouse and Doctor S. Geisser of the National Institutes of Health, for suggesting this possibility and for their advice in the construction of a robust analog of J.) Let us see what can be done along this line.

The second assumption guarantees, through the Central Limit Theorem, that all we have to do is ascertain the expected values of the mean and

variance of $S_i(d^2)$ for each subject.

Let us examine the distribution of $S_i(d^2)$ that results when the second assumption does not hold. Since the ELP assumption results in a rectangular distribution where each possible rank-order has a probability of $\frac{1}{n!}$ the effect of deviations from this assumption will be to make the probabilities for the various rank-orders unequal. If all the probability is concentrated in one rank-order, the variance of $S_i(d^2)$ will, of course, be zero.

Let us assume that the distribution of probabilities is such that the expected mean $S_i(d^2)$ remains $\frac{1}{6}(n^3-n)$, i.e., there is no trend. The maximum variance for $S_i(d^2)$ occurs when all the probabilities are concentrated equally on the minimum and maximum values of $S_i(d^2)$, zero and $\frac{1}{3}(n^3-n)$. (I am indebted to S. Geisser for a proof of this.) Then the variance associated with each subject is $\frac{1}{6}(n^3-n)^2$.

If we have m subjects whose mean $S_i(d^2)$, on the null hypothesis, is $(n^3-n)/6$, then the expected mean of J remains $m(n^3-n)/6$ but the maximum variance of J is $m(n^3-n)^2/6$.

Then

$$(8) \qquad g = \frac{J-\frac{m}{6}(n^3-n)+1}{\frac{1}{6}(n^3-n)\sqrt{m}}$$

represents a test of the assumption that the average rank-order correlation with the a priori rank-order is zero for each subject where the n treatment distributions may have any shape or any kind of dependence on one another. It is, however, an extremely weak test and requires many more subjects than J does for significance.

What are the alternative hypotheses against which g, the robust analog of the J test, should be used? This is difficult to answer exactly. Suppose we have n treatments each of which gives rise to a symmetric distribution. When the distributions are identical, with constant dependency, the J test will be appropriate. But this is almost equal to the conditions that must be met for analysis-of-variance. When the distributions are <u>not</u> identical, when the dependencies are <u>not</u> constant, then the g test will be safe.

It is possible, as we have mentioned previously, to have n treatment distributions such that the population means are equal but the expected value of Spearman's rho is non-zero. In other words, neither J nor its robust analog, g, is primarily a test for the equality of the n correlated means. They are tests for trend which weight each subject's contribution equally.

Another anomalous situation for the J test would be the case where there are two or more distinct groups of subjects and the n treatments may be expected to exhibit a different trend in each group. (This is analogous to a significant block-by-treatment interaction in the two-way analysis-of-variance.) As long as the expected value of Spearman's rho is positive for every subject, the robust test, g, will be appropriate. However, if the expected values are positive and negative, I know of no single simple test of significance.

Summary

A statistic, J, has been proposed as a test for the differences between a set of n correlated means when m subjects have been subjected → next page

to n treatments. The null hypothesis is that all permutations of the n scores for each subject are equally likely. The alternative hypothesis is that there is a trend which can be specified by the experimenter in the form of a theoretical rank-order. The J statistic is the sum of the $S(d^2)$ values that can be computed between the hypothetical rank-order and the observed rank-order for each of the m subjects. It is therefore equivalent to an average Spearman rank-order correlation. There is no parametric test that can be substituted for the non-parametric trend test.

The distribution of J has been shown to be nearly normal for the cases where $mn > 12$, and distribution tables have been given for those cases where the normal curve does not give an adequate fit.

The J test is essentially identical with Jonckheere's P test for trend except that it is based on Spearman's rho rather than Kendall's tau. It is suggested that the greater sensitivity of rho usually leads to equal or greater power for J as compared to P.

The basic assumption, that all permutations of a subject's set of n scores are equally likely, implies severe restrictions on the n-variate distribution of raw scores. These restrictions are almost equivalent to those for analysis of variance. A preliminary test is suggested to indicate when the significance of J may be due to deviations from the "equally likely permutations" assumption, rather than to differences between the n correlated means. A robust (but weak) analog of J is proposed for those cases where this assumption does not hold.

Insert Table 7 here

References

1. H. Hotelling (1931). The generalization of students' ratio. Ann. Math. Statist., 2, 359-378.

2. A. R. Jonckheere (1954a). A distribution-free k-sample test against ordered alternatives. Biometrika, 41, 133-145.

3. A. R. Jonckheere (1954b). A test of significance for the relation between m rankings and k ranked categories. Brit. J. Statist. Psychol., 7, 93-100.

4. M. G. Kendall (1948). Rank correlation methods. Griffin and Co., London.

5. E. L. Lehmann and C. Stein (1949). On the theory of some non-parametric hypotheses. Ann. Math. Statist., 20, 28-45.

6. E. S. Pearson and H. O. Hartley (1954). Biometrika tables for statisticians. (Vol. I) Cambridge University Press, Cambridge.

Table 1

Number of Errors Made in Successive Learning Trials

	Trial						
Subject	7	6	5	4	3	2	1
1	0	4	10	8	5	3	12
2	2	0	7	6	8	4	5
3	1	2	9	8	5	4	10
4	0	7	5	10	6	11	9
5	4	3	8	2	9	5	7
Total	7	16	39	34	33	27	43

Table 2

Jonckheere's Example: Rank-Order of Number of Errors Made in Each Trial of a Learning Experiment

Theoretical Rank-Order	Trial 1	2	3	4	5	6	7	$S_i(d^2)$	P_i
Subject									
1	1	3	6	5	4	2	7	28	14
2	2	1	6	5	7	3	4	34	13
3	1	2	6	5	4	3	7	20	15
4	1	4	2	6	3	7	5	18	16
5	3	2	6	1	7	4	5	34	13
							Sum	134	71

$$K = 1 - \frac{6J}{n(n^3-n)} = 1 - \frac{6(134)}{5(343-7)} = .48$$

Table 3

Seconds of Alpha during Sleep Loss

	Days of Sleep Loss			
Subjects	1	2	3	4
A	10	12	5	9
B	5	11	3	4
C	8	9	5	6
Total	23	32	13	19

Table 4

Rank-Order of Seconds of Alpha during Sleep Loss

Theoretical Rank-Order	Days 4	3	2	1	$S_i(d^2)$	P_i
Subject						
A	3	4	1	2	4	4
B	3	4	1	2	4	4
C	3	4	1	2	4	4
				Sum	12	12

$$K = 1 = \frac{6J}{m(n^3-n)} = 1 - \frac{6(12)}{3(4^3-4)} = .60$$

Table 5

Jonckheere's Example. Deviate Scores

	Trial						
Subject	7	6	5	4	3	2	1
1	-1.4	.8	2.2	1.2	-1.6	-2.4	3.4
2	.6	-3.2	-.8	-.8	1.4	-1.4	-3.6
3	-.4	-1.2	1.2	1.2	-1.6	-1.4	1.4
4	-1.4	3.8	-2.8	3.2	-.6	5.6	.4
5	2.6	-.2	.2	-4.8	2.4	-.4	-1.6

Table 6

Jonckheere's Example. Rank-Order for Deviate Scores

	Trial							
	7	6	5	4	3	2	1	
	Theoretical Rank-Order							
Subject	1	2	3	4	5	6	7	$S(d^2)$
1	3	4	6	5	2	1	7	52
2	6	2	4	5	7	3	1	76
3	4	3	5	6	1	2	7	50
4	2	6	1	5	3	7	4	36
5	7	4	5	1	6	3	2	88
m = 5	n = 7							J = 302

$$z = \frac{J-u+1}{\sigma} = \frac{302-280+1}{51.121} = .450$$

Table 7

Cumulative Probability Distribution of J

		n = 3			n = 4		n = 5	n = 6
J	n = 2	n = 3	n = 4	n = 2	n = 3	n = 2	n = 2
0	.028	.005	.001	.002	.000	.000	.000
2	.139	.032	.007	.012	.001	.001	.000
4	.250	.088	.025	.031	.003	.002	.000
6	.361	.153	.056	.056	.007	.005	.000
8	.639	.278	.109	.106	.015	.010	.001
10	.750	.444	.201	.148	.028	.017	.001
12	.856	.555	.306	.210	.045	.026	.002
14	.972	.722	.434	.281	.070	.038	.003
16	1.000	.847	.577	.366	.102	.055	.005
18		.912	.701	.443	.143	.073	.007
20		.968	.799	.557	.191	.096	.010
22		.995	.893	.634	.253	.123	.014
24		1.000	.944	.720	.315	.156	.019
26			.975	.790	.390	.190	.024
28			.993	.852	.462	.231	.031
30			.997	.894	.538	.272	.039
32			1.000	.944	.610	.320	.048
34				.968	.685	.368	.059
36				.988	.747	.421	.071
38				.998	.809	.471	.085
40				1.000	.857	.529	.101
42					.898	.579	.118
44					.930	.632	.137
46					.955	.680	.158
48					.972	.728	.181
50					.985	.769	.205
52					.993	.810	.231
54					.997	.844	.257
56					.999	.877	.287
58					1.000	.904	.318
60					1.000	.927	.349
62						.946	.381
64						.962	.415
66						.974	.448
68						.983	.482
70						.990	.518
						*	*

*The remaining probabilities are omitted because of lack of space but can be obtained by symmetry.